Assignment - 1
Machine Learning

Submitted by -
Akshay Tambe (apt 321)
Snahil Singh (ss11381)

Q1).

A	B
$\mu_1 = 9.2$	$\mu_2 = 9.6$
$\sigma_1 = 1.6$	$\sigma_2 = 1.2$

a) Given $P(B) = 4P(A)$ and we know that $P(A) + P(B) = 1$

$\therefore$ $P(A) = 1/5$, $P(B) = 4/5$

Prob. that it was from location $A = 1/5$

Prob. that it was from location $B = 4/5$

b) $x_1 = 9.3$, $x_2 = 8.8$, $x_3 = 9.8$

$C_1 : x_1, x_2, x_3 \longrightarrow A$ (location A)

$C_2 : x_1, x_2, x_3 \longrightarrow B$ (location B)

Using Baye's Rule,

$$P(C_1|x) = \frac{P(C_1) \cdot P(x|C_1)}{P(x)}$$

$P(C_1)$: Prior (derived from above question) $= 1/5$

$P(x|C_1)$: likelihood

$P(x = x_1, x_2, x_3 \mid C_1 = A) = P(x_1|A) * P(x_2|A) * P(x_3|A)$

As location A rocks are distributed according to a Gaussian distribution, we'll write pdf for gaussian distribution formula -

$$P(x_1|A) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}}$$

$$\therefore P(x_1, x_2, x_3|A) = \left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^3 e^{\frac{-1}{2\sigma_1^2}\left[(x_1-\mu_1)^2 + (x_2-\mu_1)^2 + (x_3-\mu_1)^2\right]}$$

$$= \left(\frac{1}{\sqrt{2\pi}\cdot(1.6)}\right)^3 \cdot e^{\frac{-1}{2(1.6)^2}\left[(9.3-9.2)^2+(8.8-9.2)^2+(9.8-9.2)^2\right]}$$

$$= (0.0155)(e^{-0.1035}) = 0.01397$$

$$P(x_1, x_2, x_3 \mid B) = \left(\frac{1}{\sqrt{2\pi}\,\sigma_2}\right)^3 e^{\frac{-1}{2\sigma_2^2}\left[(x_1-\mu_2)^2+(x_2-\mu_2)^2+(x_3-\mu_2)^2\right]}$$

$$= \left(\frac{1}{\sqrt{2\pi}\,(1.2)}\right)^3 e^{\frac{-1}{2(1.2)^2}\left[(9.3-9.6)^2+(8.8-9.6)^2+(9.8-9.6)^2\right]}$$

$$= 0.02812$$

$$P(x) = P(C_1)\cdot P(x \mid C_1) + P(C_2)\cdot P(x \mid C_2)$$

$$= \frac{1}{5}(0.01397) + \frac{4}{5}(0.02812) = 0.02529$$

$$\therefore \; P(C_1 \mid x) = \frac{P(C_1)\cdot P(x \mid C_1)}{P(x)} = \frac{1/5\,(0.01397)}{0.02529} = 0.11047$$

Thus, probability (posterior) that rocks are from location A is 0.11047.

c) ML hypothesis, $\underset{c}{\operatorname{argmax}}\, P(x \mid c)$ i.e either $P(x \mid C_1)$ or $P(x \mid C_2)$ (whichever is max.)

$P(x \mid C_1) = 0.01397$

$P(x \mid C_2) = 0.02812$

$\therefore$ ML hypothesis is $P(x \mid C_2) = 0.02812$, where C_2 is that rocks are from location B.

Q2) Screening 1

False +ve = 0.15 → P (positive | No disease) = 0.15

P (negative | No disease) = 0.85

false -ve = 0.10 → P (negative | disease) = 0.1

P (positive | disease) = 0.9

Screening 2,

false +ve = 0.05 → P (positive | no disease) = 0.05

P (negative | no disease) = 0.95

false -ve = 0.03 → P (negative | disease) = 0.03

P (positive | disease) = 0.97

P (disease) = 0.0002

P (no disease) = 0.9998

a) $X = \{positive\}_{S_1}$

Compare P(disease |x) and P(no disease |x)

$$P(\text{disease} \mid x) = \frac{P(\text{disease})}{P(x)} \cdot P(x \mid \text{disease}) \quad \text{— (1)}$$

$$P(\text{No disease} \mid x) = \frac{P(\text{No disease})}{P(x)} \cdot P(x \mid \text{No disease}) \quad \text{— (2)}$$

Comparing numerators from eq (1) & (2),

$$P(\text{disease} \mid x) = \frac{P(\text{disease})}{P(x)} P(x \mid \text{disease})$$

$$= 0.0002 \times 0.9 = 0.00018$$

$$P(\text{No disease} \mid X) = \frac{P(\text{No disease})}{P(X)} (P\, x \mid \text{No disease})$$

$$= 0.9998 \times 0.15 = 0.14997 \quad \text{— (4)}$$

So, P (No disease |X) > P (disease |X)

∴ MAP hypothesis is - person doesn't have a disease (Bayesian approach)

b) To find ML hypothesis,

$P(\text{disease} \mid x)_{s_1} = 0.9$

$P(\text{No disease} \mid x)_{s_1} = 0.15$

$\therefore$ ML hypothesis is $\rightarrow$ Person has a disease.

c) $x = \{\text{positive, positive}\}_{s_1, s_2}$

Posterior probability that person has a disease?

$$P(\text{Disease} \mid \underset{s_1}{\text{positive}}, \underset{s_2}{\text{positive}}) = \frac{P(\text{Disease}) \cdot P(+ve \mid D)_{s_1} \cdot P(+ve \mid D)_{s_2}}{P(+ve, +ve)_{s_1, s_2}}$$

$$P(D \mid S_1 S_2) = \frac{P(D) \cdot P(S_1 S_2 \mid D)}{P(S_1 S_2)}$$

$$P(S_1 S_2) = P(D) \cdot P(+ve_{S_1}, +ve_{S_2} \mid D) + P(ND) \cdot P(+ve_{S_1}, +ve_{S_2} \mid ND)$$
$$= P(D) \cdot P(S_1 \mid D) \cdot P(S_2 \mid D) + P(ND) \cdot P(S_1 \mid ND) \cdot P(S_2 \mid ND)$$
$$= (0.0002)(0.9)(0.97) + (0.9998)(0.15)(0.05)$$
$$= 0.007673$$

$$P(D \mid S_1 S_2) = \frac{0.0002 \times 0.9 \times 0.97}{0.007673} = 0.0228$$

Posterior prob. that a person has disease is 0.0228

Q3. $P(H) = \theta$, $P(T) = 1 - \theta$

We get TTHTTH

a) $P(TTHTTH) = (1-\theta)^4 \theta^2 \quad \{= P(T) \cdot P(T) \cdot P(H) \cdot P(T) \cdot P(T) \cdot P(H)\}$

b) $\log P(TTHTTH) = \log \{(1-\theta)^4 \theta^2\} = 4 \log(1-\theta) + 2 \log \theta$

c) ML estimate of $\theta = \underset{\theta}{\text{argmax}}\ P(TTHTTH \mid \theta)$; need to compute $\underset{\theta}{\text{argmax}} \log P(TTHTTH \mid \theta)$

Partial derivative of $\log P(TTHTTH)$ w.r.t θ shd be zero.

$$\frac{\partial \log P(TTHTTH)}{\partial \theta} = \frac{\partial(4\log(1-\theta) + 2\log\theta)}{\partial \theta} = 4\frac{\partial \log(1-\theta)}{\partial(1-\theta)} \cdot \frac{\partial(1-\theta)}{\partial\theta} + 2\frac{\partial \log\theta}{\partial\theta}$$

$$= 4(\cancel{-\theta}) + \quad = \cancel{2}/\frac{4}{\theta}$$

$$\Rightarrow \frac{2}{\theta} - \frac{4}{1-\theta} = 0$$

$$\Rightarrow 2(1-\theta) - 4\theta = 0$$

$$\Rightarrow 2 - 6\theta = 0 \Rightarrow \theta = 1/3$$

ML estimate is at $\theta = 1/3$.

Q4.

a) $S=3$, $X = \{3, 1, 1, 2, 3\}$

$\theta_1 = \theta_v \therefore v = 1$

$\therefore \frac{t}{N} = \frac{2}{5}$ is ML estimate.

b) ~~9A / 21~~, Here, $m=1$, $\therefore$ ML estimate $\frac{t+1}{N+S} = \frac{2+1}{5+3} = \frac{3}{8} = 0.375$

c) $S=2$ (possible values of x)

$N \to$ # values in set

$t \to$ # occurances of 1

$\theta_1 = \theta_v \Rightarrow v = 1$

Given N independent occurances where possible values of x can be only 2 $\Rightarrow P(x|\theta) = \theta^{t}(1-\theta)^{N-t} \therefore P(X|\theta_1) = \theta_1^{t}(1-\theta_1)^{N-t}$

Given $\to P(\theta_1) = 6\theta_1(1-\theta_1)$, $\theta_{map} = \arg\max_\theta P(\theta_1) \cdot P(X|\theta_1)$

$= \arg\max_\theta 6\theta_1(1-\theta_1)\theta_1^t(1-\theta_1)^{N-t} = \arg\max [\log 6 + (t+1)\log\theta_1 + (N-t+1)\log(1-\theta_1)]$

Diffentiating w.r.t θ

$\frac{d}{d\theta} \Rightarrow \frac{t+1}{\theta_1} - \frac{(N-t+1)}{(1-\theta_1)} \quad \therefore \frac{t+1}{\theta_1} = \frac{N-t+1}{1-\theta_1} \Rightarrow \theta_1 = \frac{t+1}{N+2}$

Also, given $S=2$, comparing with estimate (2), $S=2$ & $m=1$ (add-1-smoothing)

θ_1 (estimate ii) $= \frac{t+1}{N+S}$

Also, true for m smoothing ($m \in [0,1]$), θ_1(estimate iii) $= \frac{t+m}{N+Sm}$

Solution 5

Classes : $C_1 = +$

$C_2 = -$

$x_1 \rightarrow \{Low, Medium, High\}$.

$x_2 \rightarrow \{Yes, No\}$.

$x_3 \rightarrow \{Red, Green\}$

a) Add m-smoothing is given by formula,

$$P(x|c) = \frac{count(x,c) + m}{count(c) + (m \times \text{possible values of } x)}$$

Using above, (Given $m = 0.3$).

$$P(x_1 = Low | +) = \frac{1 + 0.3}{2 + (0.3 \times 3)} = 0.4483$$

$count(Low, +) = 1$
$count(c)_+ = 2$.
$count(x_1) = 3$.

$$P(x_2 = Yes | +) = \frac{0 + 0.3}{2 + (0.3 \times 2)} = 0.1154$$

$count(Yes, +) = 0$
$count(c)_+ = 2$
$count(x_2) = 2$

$$P(x_3 = Green | +) = \frac{1 + 0.3}{2 + (0.3 \times 2)} = 0.5$$

$count(Green, +) = 1$
$count(c)_+ = 2$
$count(x_3) = 2$.

Q5. continued ...

$P(x_1 = \text{low} \mid -) = \dfrac{1+0.3}{3+(0.3\times 3)} = 0.3333$

count (low, −) = 1
count (c)₋ = 2
count (x_1) = 3

$P(x_2 = \text{yes} \mid -) = \dfrac{2+0.3}{3+(0.3\times 2)} = 0.6389$

count (yes, −) = 2
count (c)₋ = 3
count (x_2) = 2

$P(x_3 = \text{Green} \mid -) = \dfrac{2+0.3}{3+(0.3\times 2)} = 0.6389$

count (Green, −) = 2
count (C)₋ = 3
count (x_3) = 2

b) Comparing $P(x \mid +)$ and $P(x \mid -)$

where, $x = \{\text{low, yes, green}\}$, $P(+) = \frac{2}{5}$, $P(-) = \frac{3}{5}$

By using Naive-Bayes assumption

$P(x \mid c) = P(x_1 = \text{low} \mid c) * P(x_2 = \text{yes} \mid c) * P(x_3 = \text{green} \mid c)$

$P(x \mid +) = 0.4483 \times 0.1154 \times 0.5 = 0.0259$

$P(x \mid -) = 0.3333 \times 0.6389 \times 0.6389 = 0.13605$

c) ML hypothesis

From b) we can say $P(x \mid -) > P(x \mid +)$

∴ ML label for x {low, Yes, green} is −

d) MAP hypothesis

$P(+) = 2/5$, $P(-) = 3/5$

$P(x \mid c) = P(c) \cdot P(x_1 = \text{low} \mid c) * P(x_2 = \text{yes} \mid c) * P(x_3 = \text{green} \mid c)$

$P(x \mid +) = 2/5 \times 0.0259 = 0.01036$

$P(x \mid -) = \cancel{3/5 \times 0.1809} = 3/5 \times 0.13605 = 0.08163$

$P(x \mid -) > P(x \mid +)$, MAP hypothesis $\rightarrow P(x \mid -)$

MAP label for x {low, yes, green} is −